Incentives and Machine Learning (CSC 482A/581A)
Lectures 19–21

Nishant Mehta

1 Correlated Equilibria and Coarse Correlated Equilibria

Let each player $i \in [n]$ have strategy set A_i, and define $A = A_1 \times \cdots \times A_n$. We use $p \in \Delta(A)$ to refer to a joint probability distribution over strategy profiles. In the study of Nash equilibria, we only considered product distributions (so players draw their strategies independently). We now allow for general distributions p; the players' strategies can be dependent.

1.1 Correlated Equilibria

We begin by introducing the notion of correlated equilibria (CE).

Definition 1. We say $p \in \Delta(A)$ is correlated equilibrium if for all agents $i \in [n]$, for all strategies $a_i, a'_i \in A_i$, it holds that

$$\mathsf{E}_{a \sim p}[c_i(a_i, a_{-i}) \mid a_i] \leq \mathsf{E}_{a \sim p}[c_i(a'_i, a_{-i}) \mid a_i].$$

Note that on the LHS, the expectation is really only over a_{-i} since we condition on a_i.[1]

To better understand the definition, consider the following game protocol that involves a coordinator. The coordinator has a joint probability distribution p over strategy profiles, and this distribution is known to all agents.

1. First, the coordinator draws a strategy profile a from p.
2. Next, the coordinator tells each player i the strategy a_i (but provides this player no other information about a).
3. Finally, each player is free to play whatever strategy she wants.

In the above protocol, it's useful to think of a_i as the strategy that the coordinator recommends to player i. In the real-world, like in the above protocol, people don't have to follow recommendations.

Coming back to the definition of correlated equilibria, we say that p is a correlated equilibrium if, conditional on being recommended a_i, player i's expected cost for playing any other strategy cannot be smaller. Thus, the agent has no incentive to deviate from playing her recommended strategy a_i.

To appreciate the richness of the above protocol compared to the situation where there was no coordinator, let's look at a game called the traffic light game. Each agent has the same two actions: Go and Stop. If both agents select Stop, then then they each have a cost of 1. If one agent selects Stop and the other selects Go, then the agent that played Go has a cost of 0, while the agent that

[1] A clearer definition would be: for all $a''_i, a'_i \in A_i$, it holds that $\mathsf{E}_{a \sim p}[c_i(a_i, a_{-i}) \mid a_i = a''_i] \leq \mathsf{E}_{a \sim p}[c_i(a'_i, a_{-i}) \mid a_i]$; we opted for the presentation above because it is standard in the literature, so it's good to get used to it.

played Stop has a cost of 1. Finally, if both agents play Go, then they both suffer a terribly high cost of 100 (car accidents are no fun). This game is expressed in the cost table belowgame table below.

		Player 2: Go	Player 2: Stop
Player 1	Go	$(100, 100)$	$(0, 1)$
	Stop	$(1, 0)$	$(1, 1)$

Suppose the agents play this game without a coordinator. This puts us in the protocol that we have considered in the course up until this lecture. What are the pure strategy Nash equilibria, if any? It is easy to see that there are two of them, corresponding to (Go, Stop) and (Stop, Go). This is OK, but it's not great if the same agent is always the one that plays Stop (this agent will always pay a cost). This asymmetry seems undesirable.

Let's try to seek a better equilibrium by considering mixed strategy Nash equlibria (MNE). To identify an MNE, suppose that Player 1 adopts the mixed strategy that plays Go with probability p and Stop with probability $1-p$. Then Player 2's expected cost for playing Go is $p \cdot 100 + (1-p) \cdot 0 = 100p$, and her expected cost for playing Stop is $p \cdot 1 + (1-p) \cdot 1 = 1$. Note that p is selected by Player 1 such that these two costs are equal, then any mixed strategy $(q, 1-q)$ adopted by Player 2 is a best response to Player 1. Solving for p, we see that Player 2 is indifferent between her strategies when $p = \frac{1}{100}$. Now, if Player 2 also sets $q = p = \frac{1}{100}$, then Player 1 is best responding to Player 2 by the same reasoning. Hence, when both players use the mixed strategy $(0.01, 0.99)$, they are in an MNE.

Let us reflect a bit about this MNE. Qualitatively, it has some bad properties. First, most of the time, neither player is going anywhere. Indeed, with probability $0.99^2 \approx 0.98$, both players both select Stop. Moreover, with some small (but non-zero) probability $0.01^2 = 0.0001$, there is a car accident. Such a situation would be completely avoidable if the game protocol allowed for a coordinator (via a traffic light). Finally, let's compute each player's expected cost: $p^2 \cdot 100 + p(1-p) \cdot 0 + (1-p)p \cdot 1 + (1-p)^2 \cdot 1 = 100p + (1-p) = 0.01 + 0.99 = 1$. Ignoring the issue of the chance of a car accident, this is not terrible; however, its social cost is 2, whereas for either of the PNE the social cost is 1. The issue with the MNE is that that concept of Nash equilibria requires that player's independently draw their strategies. That is, the joint distribution over strategy profiles must be a product distribution. What we would like to do is have a joint distribution over strategy profiles that puts equal probability mass on the strategy profiles (Go, Stop) and (Stop, Go). Unfortunately, this simply is not possible for product distributions.

There's a much better solution that puts the nail in the coffin of this MNE, allows for symmetry between the players (unlike the two PNE), and has the same social cost as the PNE. For this, we extend the protocol to allow for a coordinator exactly as we discussed earlier. The coordinator can draw strategy profiles from a distribution p (we now once again use p to refer to the joint distribution over strategy profiles). Suppose the coordinator puts equal probability mass on (Stop, Go) and (Go, Stop). Now, if the coordinator draws a strategy profile $a \sim p$ and a_1 is Stop, the coordinator suggests to Player 1 to play Stop. Player 1, knowing the distribution p but not a_2, can reason that the other player was told to Go. Therefore, Player 1 should not deviate and should play the recommended strategy Stop. From a symmetric reasoning, the player (who is recommended to play Go) should indeed play Go. Hence, this p is a correlated equilibrium. Moreover, a basic computation shows that each player has expected cost 0.5 under p.

1.2 Coarse Correlated Equilibria

Correlated equlibria can be relaxed to coarse correlated equilibria (CCE).

Definition 2. We say $p \in \Delta(A)$ is coarse correlated equilibrium if for all agents $i \in [n]$, for all strategies $a_i' \in A_i$, it holds that

$$\mathsf{E}_{a \sim p}[c_i(a_i, a_{-i})] \leq \mathsf{E}_{a \sim p}[c_i(a_i', a_{-i})].$$

Note that the definition is the same as correlated equilibrium except that we dropped the conditioning. What does it mean to not condition? Well, it means that the agent must decide whether to deviate *before* the coordinator suggests to the agent what action to play. This might be hard to justify in general; once the coordinator reveals to the agent the suggested action, the agent would have to be trusted to not use this new information when it decides whether to deviate.

2 No-external-regret and approximate coarse correlated equilibria

Previously, we saw how no-regret dynamics can be used to obtain approximate Nash equilibria for two-player zero-sum games. The idea was for each player to run a no-external-regret[2] algorithm like Hedge. Then, the strategy profile formed by taking the time-average of each player's mixed strategy gives an approximate Nash equilibrium. It turns out that precisely the same version of no-regret dynamics also gives an approximate coarse correlated equilibrium for finite, n-player games. Moreover, there is a variant of no-regret dynamics (based on a new notion of regret) that also can be used to approximate correlated equilibria. We will see this new notion of regret, called "swap regret", a little later. For now, let's see how to approximate coarse correlated equilibria using concepts that we've already seen.

Throughout this lecture, we will put the round index in the superscript so that, whenever needed, the subscript can be used to indicate the player. This notational choice is more standard in the literature. Let $\ell_i^{(1)}, \ldots, \ell_i^{(T)}$ be a sequence of cost functions for player i, where each loss function $\ell_i^{(t)}$ is a map $\ell_i^{(t)} \colon A_i \to [0,1]$. We say a learning algorithm that plays mixed strategies $p_i^{(1)}, \ldots, p_i^{(T)}$ is no-external-regret if there exists[3] $\varepsilon_T = o(1)$ such that for all $a_i' \in A_i$,

$$\frac{1}{T}\left(\sum_{t=1}^{T} \mathsf{E}_{a_i^{(t)} \sim p_i^{(t)}}[\ell_i^{(t)}(a_i^{(t)})] - \sum_{t=1}^{T} \ell_i^{(t)}(a_i')\right) \leq \varepsilon_T.$$

To use this definition in the context of player i playing an n-agent repeated game, we set $\ell_i^{(t)}(a_i^{(t)}) = \mathsf{E}_{a_{-i}^{(t)} \sim p_{-i}^{(t)}}\left[c_i(a_i^{(t)}, a_{-i}^{(t)})\right]$, where $p_{-i}^{(t)}$ is the distribution (induced by $p^{(t)}$) over actions of all players but player i.

Before presenting our result for approximating CCE using no-regret dynamics, we first need to formally define approximate CCE.

Definition 3. We say $p \in \Delta(A)$ is an ε-approximate coarse correlated equilibrium if for all agents $i \in [n]$, for all strategies $a_i' \in A_i$, it holds that

$$\mathsf{E}_{a \sim p}[c_i(a_i, a_{-i})] \leq \mathsf{E}_{a \sim p}[c_i(a_i', a_{-i})] + \varepsilon.$$

[2] For precision, we now call the term "no-regret" from a previous lecture as "no-external-regret".

[3] Recall that $\varepsilon_T = o(1)$ if $\varepsilon_T \to 0$ as $T \to \infty$.

Theorem 1. *Suppose each player i uses a no-external-regret algorithm whose worst-case average regret is at most ε_T. For each round t, define $p^{(t)} = p_1^{(t)} \times \cdots \times p_n^{(t)}$. Let $\bar{p} = \frac{1}{T}\sum_{t=1}^{T} p^{(t)}$. Then $\bar{p}$ is an ε_T-approximate coarse correlated equilibrium.*

Proof. For any player i and strategy $a_i' \in A_i$, observe that

$$\begin{aligned}
&\mathsf{E}_{a\sim\bar{p}}[c_i(a)] - \mathsf{E}_{a\sim\bar{p}}[c_i(a_i', a_{-i})] \\
&= \frac{1}{T}\sum_{t=1}^{T} \mathsf{E}_{a\sim p^{(t)}}[c_i(a)] - \frac{1}{T}\sum_{t=1}^{T} \mathsf{E}_{a\sim p^{(t)}}[c_i(a_i', a_{-i})] \\
&= \frac{1}{T}\sum_{t=1}^{T} \mathsf{E}_{a_i^{(t)}\sim p_i^{(t)}}\left[\mathsf{E}_{a_{-i}^{(t)}\sim p_{-i}^{(t)}}\left[c_i(a_i^{(t)}, a_{-i}^{(t)})\right]\right] - \frac{1}{T}\sum_{t=1}^{T} \mathsf{E}_{a_{-i}^{(t)}\sim p_{-i}^{(t)}}\left[c_i(a_i', a_{-i}^{(t)})\right] \\
&= \frac{1}{T}\left(\sum_{t=1}^{T} \mathsf{E}_{a_i^{(t)}\sim p_i^{(t)}}[\ell_i^{(t)}(a_i^{(t)})] - \sum_{t=1}^{T} \ell_i^{(t)}(a_i')\right) \\
&\le \varepsilon_T,
\end{aligned}$$

where the last line follows since player i has no-external-regret. □

3 Swap Function Formulation of CE, and Swap Regret

When it comes to approximating correlated equilibria, the original formulation of correlated equilibria in Definition 1 is not the most convenient one to work with. There is a different, equivalent formulation of correlated equilibria which can directly be connected to the soon-to-be-defined notion of swap regret. Using this new definition, it will be easy to show how no-regret dynamics — based on swap regret — can be used to approximate correlated equilibria.

3.1 Formulation of correlated equilibria via swap functions

It is possible to show that the following definition of correlated equilibria is equivalent to Definition 1. Showing this equivalence likely will be an exercise in the second problem set.

Definition 4. We say $p \in \Delta(A)$ is a correlated equilibrium if for all agents $i \in [n]$, for all swap functions $\phi_i\colon A_i \to A_i$, it holds that

$$\mathsf{E}_{a\sim p}[c_i(a_i, a_{-i})] \le \mathsf{E}_{a\sim p}[c_i(\phi_i(a_i), a_{-i})].$$

We now straightforwardly adapt the above definition to define approximate correlated equilibria.

Definition 5. We say $p \in \Delta(A)$ is an ε-approximate correlated equilibrium if for all agents $i \in [n]$, for all swap functions $\phi_i\colon A_i \to A_i$, it holds that

$$\mathsf{E}_{a\sim p}[c_i(a_i, a_{-i})] \le \mathsf{E}_{a\sim p}[c_i(\phi_i(a_i), a_{-i})] + \varepsilon.$$

3.2 Swap regret

The notion of regret that we have seen thus far is known as external regret. One way to view external regret against some fixed action a' is for the learning algorithm to consider how its performance would have changed if, in each round t, the learning algorithm played a' instead of a_t (so, it swaps all actions for a').

Suppose we upgrade this idea as follows. The learning algorithm looks back at its history of plays $a^{(1)}, \dots a^{(T)}$. For each action $a \in A$, for all rounds where it played that action, it considers swapping the action for another action $a' \in A$. For example, if you buy a piece of fruit each day, your comparator could be a strategy which makes all of the following swaps:

- for any day when you bought an orange, buy a mango instead;
- for any day when you bought an apple, buy a pear instead;
- for any day when you bought a grapefruit, buy an orange instead.

An easy way to express these swaps is by way of a swap function $\phi\colon A \to A$ which, given a played action a, considers playing action $\phi(a)$ instead.

For the definition below, we consider an online learning algorithm that, in each round t, plays an action $a^{(t)} \in A$ and suffers a loss according to a loss function $\ell^{(t)}\colon A \to \mathbb{R}$.

Definition 6. The *swap regret* of a sequence of actions $a^{(1)}, \dots, a^{(T)}$ relative to swap function $\phi\colon A \to A$ is

$$\sum_{t=1}^{T} \ell^{(t)}(a^{(t)}) - \sum_{t=1}^{T} \ell^{(t)}(\phi(a^{(t)})).$$

4 No-swap-regret and approximate correlated equilibria

Similar to no-external-regret, we say a learning algorithm that plays mixed strategies $p_i^{(1)}, \dots, p_i^{(T)}$ is no-swap-regret if there exists $\varepsilon_T = o(1)$ such that for all swap functions $\phi\colon A_i \to A_i$,

$$\frac{1}{T}\left(\sum_{t=1}^{T} \mathsf{E}_{a_i^{(t)} \sim p_i^{(t)}}[\ell_i^{(t)}(a_i^{(t)})] - \sum_{t=1}^{T} \mathsf{E}_{a_i^{(t)} \sim p_i^{(t)}}[\ell_i^{(t)}(\phi(a_i^{(t)}))]\right) \le \varepsilon_T.$$

Theorem 2. *Suppose each player i uses a no-swap-regret algorithm whose worst-case average regret is at most ε_T. For each round t, define $p^{(t)} = p_1^{(t)} \times \cdots \times p_n^{(t)}$. Let $\bar{p} = \frac{1}{T}\sum_{t=1}^{T} p^{(t)}$. Then $\bar{p}$ is an ε_T-approximate correlated equilibrium.*

Proof. For any player i and swap function $\phi\colon A_i \to A_i$, observe that

$$\begin{aligned}
&\mathsf{E}_{a \sim \bar{p}}[c_i(a)] - \mathsf{E}_{a \sim \bar{p}}[c_i(\phi(a_i), a_{-i})] \\
&= \frac{1}{T}\sum_{t=1}^{T} \mathsf{E}_{a \sim p^{(t)}}[c_i(a)] - \frac{1}{T}\sum_{t=1}^{T} \mathsf{E}_{a \sim p^{(t)}}[c_i(\phi(a_i), a_{-i})] \\
&= \frac{1}{T}\sum_{t=1}^{T} \mathsf{E}_{a_i^{(t)} \sim p_i^{(t)}}\left[\mathsf{E}_{a_{-i}^{(t)} \sim p_{-i}^{(t)}}\left[c_i(a_i^{(t)}, a_{-i}^{(t)})\right]\right] - \frac{1}{T}\sum_{t=1}^{T} \mathsf{E}_{a_i^{(t)} \sim p_i^{(t)}}\left[\mathsf{E}_{a_{-i}^{(t)} \sim p_{-i}^{(t)}}\left[c_i(\phi(a_i^{(t)}), a_{-i}^{(t)})\right]\right] \\
&= \frac{1}{T}\left(\sum_{t=1}^{T} \mathsf{E}_{a_i^{(t)} \sim p_i^{(t)}}[\ell_i^{(t)}(a_i^{(t)})] - \sum_{t=1}^{T} \mathsf{E}_{a_i^{(t)} \sim p_i^{(t)}}\left[\ell_i^{(t)}(\phi(a_i^{(t)})\right)\right] \\
&\le \varepsilon_T,
\end{aligned}$$

where the last line follows since player i has no-swap-regret.

$\square$